The SuperCOSMOS all-sky galaxy catalogue

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **462,** 2085–2098 (2016)
Advance Access publication 2016 July 30

doi:10.1093/mnras/stw1818

The SuperCOSMOS all-sky galaxy catalogue

J. A. Peacock,[1]★ N. C. Hambly,[1] M. Bilicki,[2] H. T. MacGillivray,[1] L. Miller,[3] M. A. Read[1] and S. B. Tritton[1]

[1]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[2]*Sterrewacht Leiden, Universiteit Leiden, Niels Bohrweg 2, NL-2333 CA Leiden, the Netherlands*
[3]*Department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*

Accepted 2016 July 21. Received 2016 June 16

ABSTRACT

We describe the construction of an all-sky galaxy catalogue, using SuperCOSMOS scans of Schmidt photographic plates from the UK Schmidt Telescope and Second Palomar Observatory Sky Survey. The photographic photometry is calibrated using Sloan Digital Sky Survey data, with results that are linear to 2 per cent or better. All-sky photometric uniformity is achieved by matching plate overlaps and also by requiring homogeneity in optical-to-2MASS colours, yielding zero-points that are uniform to 0.03 mag or better. The typical AB depths achieved are $B_J < 21$, $R_F < 19.5$ and $I_N < 18.5$, with little difference between hemispheres. In practice, the I_N plates are shallower than the B_J and R_F plates, so for most purposes we advocate the use of a catalogue selected in these two latter bands. At high Galactic latitudes, this catalogue is approximately 90 per cent complete with 5 per cent stellar contamination; we quantify how the quality degrades towards the Galactic plane. At low latitudes, there are many spurious galaxy candidates resulting from stellar blends: these approximately match the surface density of true galaxies at $|b| = 30^\circ$. Above this latitude, the catalogue limited in B_J and R_F contains in total about 20 million galaxy candidates, of which 75 per cent are real. This contamination can be removed, and the sky coverage extended, by matching with additional data sets. This SuperCOSMOS catalogue has been matched with 2MASS and with *WISE*, yielding quasi-all-sky samples of respectively 1.5 million and 18.5 million galaxies, to median redshifts of 0.08 and 0.20. This legacy data set thus continues to offer a valuable resource for large-angle cosmological investigations.

Key words: methods: observational – techniques: photometric – catalogues – surveys – galaxies: photometry.

1 INTRODUCTION

Large galaxy catalogues are an essential tool for any cosmological study that aims to inspect the large-scale distribution of matter in the universe. Although the microwave background gives a purer probe of cosmological deviations from homogeneity, the pattern of galaxy clustering is the most direct and spectacular manifestation of these density fluctuations. Historically, statistical studies of this clustering have contributed hugely to the establishment of the current flat vacuum-dominated cosmological standard model (e.g. Efstathiou, Sutherland & Maddox 1990; Efstathiou et al. 2002), which has since been confirmed in many different ways (e.g. Astier & Pain 2012; Aubourg et al. 2015; Planck Collaboration XIII 2015).

The older generation of work in this area was dominated by photographic plates; for many years, digital detectors were incapable of surveying the required areas of sky. The most influential photographic catalogue was the APM survey (Maddox et al. 1990a; Maddox, Efstathiou & Sutherland 1990b), which was based on scans of the Southern hemisphere UK Schmidt blue survey plates. This contained about 20 million galaxies over 4300 deg^2 to blue magnitudes as faint as 22, and formed the original input catalogue for the highly successful 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2001, 2003), and (in part) for the 6dF Galaxy Survey (6dFGS; Jones et al. 2009). Newer generations of optical imaging surveys are however based on CCD data, and the five-band Sloan Digital Sky Survey (SDSS) sets the standard in this respect, having released data for 208 million galaxies over 31 637 deg^2 of imaging, of which over 1 per cent have spectroscopic redshifts (DR12; Alam et al. 2015).

Despite their technological obsolescence, the legacy photographic Schmidt surveys nevertheless retain one key advantage: they cover the whole sky. Digital surveys have already reached this stage in the infrared (2MASS, Skrutskie et al. 2006; *WISE*, Wright et al. 2010) or ultraviolet (*GALEX*; Morrissey et al. 2007), but this is not yet the case in optical wavebands. For the latter, we will

★E-mail: jap@roe.ac.uk

undoubtedly achieve this goal in due course; but in the meantime there are a variety of science applications that require such data over the full sky, and which can achieve interesting results using the existing material – whose quality turns out to be perhaps higher than hitherto suspected, in a tribute to its creators. The SuperCOSMOS measuring machine was therefore used to scan the best available photographic data and extract the full information in the plates. This process was initially carried out in the Southern hemisphere, based on the UK Schmidt Telescope (UKST; Hambly et al. 2001a,c; Hambly, Irwin & MacGillivray 2001b); this was followed by scanning of the Second Palomar Observatory Sky Survey (POSS2; Reid et al. 1991). The photographic material, hypersensitization strategy, filter set and overall sensitivity of UKST and POSS2 are broadly comparable, allowing a reasonably homogeneous coverage of the sky. The photometric calibration of both these surveys in a uniform and consistent manner is possible, thanks to the SDSS and in particular also to the all-sky coverage offered in the 1–2 μm wavebands by the Two-Micron All-Sky Survey (2MASS; e.g. Jarrett et al. 2000, 2003). Although it may seem implausible that near-infrared (near-IR) data could be used to calibrate optical data, it turns out that the statistical power of insisting on uniform optical-to-IR colours greatly aids the robustness of the calibration.

This calibration was carried out in 2007; the resulting publicly available SuperCOSMOS Science Archive (SSA) has since been curated by Edinburgh's Wide-Field Astronomy Unit (surveys.roe.ac.uk/ssa). The galaxy catalogue has been used in past work involving one of the present authors (Francis & Peacock 2010a,b). It was also employed to generate photometric redshifts for the 2MASS galaxy catalogue (2MPZ: Bilicki et al. 2014), yielding 1.5 million galaxies to a median redshift of 0.08, with a typical redshift precision of $\sigma_z \simeq 0.015$. More recently, we have extended this exercise to produce a joint optical–*WISE* all-sky galaxy catalogue (Bilicki et al. 2016); this contains 18.5 million galaxies to a median redshift of 0.20, with a typical redshift precision of $\sigma_z/(1+z) \simeq 0.033$. In addition, several other surveys have benefited from the SuperCOSMOS data, including HIPASS (Doyle et al. 2005), 6dFGS (Jones et al. 2009) or AT20G (Murphy et al. 2010).

In order to document the material used in the above work, and to assist other users of the public data, this paper describes the SuperCOSMOS calibration process in some detail. Section 2 describes the input SuperCOSMOS data; Section 3 gives colour equations to SDSS and discusses variations of photometry with position within a given plate; Section 4 discusses the achievement of all-sky uniformity with the aid of 2MASS; Section 5 discusses the completeness and reliability of the catalogue; and Section 6 sums up.

2 SUPERCOSMOS DATA

The operation of the SuperCOSMOS plate measuring machine and its application to the production of object catalogues is described by Hambly et al. (2001a,b,c). The machine itself has since been decommissioned, following the completion of its scanning programme. In brief, plates were digitized at relatively high spatial resolution (0.67 arcsec pixels) and high dynamic range (15 bit). The photographic transmission values were measured by a scanning CCD and converted to an estimate of linear intensity. Image analysis was then carried out by grouping connected pixels that lie above a threshold, from which basic image parameters could be deduced, particularly isophotal magnitude and image area.

Morphological classification of SuperCOSMOS image data is in essence based on the area of images as a function of magnitude. Each image is allocated a profile statistic, η, which is a measure of image sharpness scaled to have zero mean and unit standard deviation (see Hambly et al. 2001b for the details of how this is defined). Images for which η lies above 2.5 are classified as extended sources – i.e. galaxy candidates. This classification can be performed separately for each plate, but an improved overall classification can be derived when the catalogues from different plates are paired up. Where a given object is detected in more than one waveband, the η values are summed; the image is classified as extended if $\sum_i \eta_i/\sqrt{N_{\rm detections}} > 2.5$ (SuperCOSMOS parameter `meanClass = 1`). Where an image is classified as extended, an attempt is made to use the brightness profile within the above-threshold pixels to extrapolate a correction of the isophotal photometry to an approximate total magnitude, assuming that the brightness profile can be fitted by a Gaussian.

This method is more likely to yield clean samples of stars, since merged pairs of stars will be classed as galaxies. The SuperCOSMOS image analyser does attempt to deblend images, based on data at multiple thresholds. This is relatively successful for brighter objects, but inevitably the bulk of objects lie near the plate limit and thus have sufficient S/N for detection as extended sources, but not sufficient to diagnose image blends. Various unpublished attempts have been made over the years to use more sophisticated image classification algorithms, but without a great deal of success. As a result, a significant problem with galaxy catalogues obtained from these photographic plates is corruption by merged stellar images at low Galactic latitudes. This is also a potential issue with digital data, of course, but the modest image quality of the Schmidt plates (full width at half-maximum, FWHM $\simeq 2$ arcsec) exacerbates the problem – although the difficulty is greater still with the 6 arcsec FWHM of *WISE* (Bilicki et al. 2016; Kurcz et al. 2016).

In addition to morphological classification, each image is assigned a quality flag designed to pick out problematic images that are likely to be plate artefacts or corrupted in some way. This information is available for each plate (`qualB, qualR2, qualI`) – where 'R2' distinguishes the default red data from scans of the shallower plates from the ESO Schmidt and the first-generation Palomar Schmidt ('R1'). We reject objects where either `qualB` or `qualR2` $\geq$2048, indicating that the image probably results from a stellar halo or diffraction spike, or a satellite trail.

Catalogues of stellar and galaxy images generated in this way have been produced for photographic plates from the UKST and POSS2 Schmidt surveys. Together, these provide approximately homogeneous all-sky coverage in three wavebands: the IIIa-J, IIIa-F and IV-N plates, which we shall generally refer to here as $B_{\rm J}$, $R_{\rm F}$ and $I_{\rm N}$ surveys. This information has been available online since 2001 at http://www-wfau.roe.ac.uk/sss; the data are now also held in a relational data base at surveys.roe.ac.uk/ssa, which provides multicolour information on 1.9 billion distinct objects. It should be emphasized that the calibration of SuperCOSMOS photometry for objects classed as stellar was carried out separately (see Hambly et al. 2001b), and the current exercise was applied only to the extended images. Although it is possible to assign galaxy-style magnitudes to SuperCOSMOS objects that are not classed as extended by going back to the raw data, this was not done for the public data set. Thus, stars and galaxies in the SSA have magnitudes that refer to two slightly different photometric systems.

2.1 Plate special cases

In principle, both UKST and POSS2 consist of 894 fields, starting with number 1 at the south and north celestial poles, respectively. In practice, field 1 was never taken for POSS2, and the four fields 895–898 were obtained instead to bracket the pole.

It should be noted that, in one case, it was not possible to obtain a blue POSS2 plate for scanning; field 504 in the Northern hemisphere was therefore taken with the UKST, and the blue data for this field refer to that system. Similarly, for the POSS2 $I_{\rm N}$ survey, 25 fields were taken by UKST to complete the survey: 559, 560, 565, 567, 569, 630, 633, 636, 637, 641, 688, 701, 703, 704, 708, 709, 713, 771, 776, 777, 779, 780, 782, 784, 796.

Both surveys were carried out using grids of field centres spaced by 5° in B1950 declination, with the $\delta = 0$ band being observed by both telescopes. We took the decision to use UKST data in this case, so that POSS2 measurements are only present for $\delta_{\rm B1950} > 2.^\circ5$. Because of this survey grid, it should only be necessary to use the data in the central $5^\circ \times 5^\circ$ of the $6^\circ \times 6^\circ$ Schmidt plates. In practice, where multiple plates cover a given sky location, the plate where the data lie closest to the plate centre is adopted in the final catalogue. But the full area was scanned and stored, which had two benefits. First, the overlap aids in the calibration of relative plate zero-points, as discussed below. Secondly, the data in the outer parts of the plates permit a cure for the 'stepwedge problem'. The stepwedges were added to the plates by exposing standard lamps in an attempt to aid calibration. For UKST, these affect the data only outside $5^\circ \times 5^\circ$, but for POSS2 the effect comes closer to the plate centre. The location of the stepwedge is known and masked out, but this has the result that a small fraction of the sky covered by POSS2 contains no galaxies in publicly available merged source catalogues. We have remedied this by returning to the original scans and using data outside the central $5^\circ \times 5^\circ$ from overlapping plates (about 2 per cent of the Northern hemisphere was affected).

3 CALIBRATION WITH SDSS

3.1 Colour equations

Photographic magnitudes have traditionally been placed on a Vega system and calibrated with respect to Johnson–Cousins magnitudes (see e.g. Blair & Gilmore 1982). Today, we should consider this in the context of the largest data set of calibrating photometry, which is the SDSS; here the magnitudes are on an AB system (Fukugita et al. 1996). In practice, we used the SDSS Petrosian magnitudes as the 'truth' for each galaxy. Before the photographic magnitudes can be calibrated against such digital photometry, they must be corrected to a genuine logarithmic measure of flux – a non-trivial process, given the highly non-linear response of photographic emulsions. The initial SuperCOSMOS processing attempted to linearize the plate transmission data, but this is hard to achieve precisely, since saturation effects can vary within the scanning resolution. As a result, the SuperCOSMOS photometry is not precisely linear. This issue is discussed below in Section 3.3; it must be approached in an iterative way, since the colour equations discussed here require linearized data, but plate calibration requires the colour equations to be known. In practice, an initial plate calibration was made assuming published colour equations (discussed below), following which direct colour equations to SDSS could be deduced, allowing a new calibration; the colour equations were stable thereafter. The relation between SDSS and SuperCOSMOS magnitudes is significantly non-linear, requiring a transformation that is quadratic in SDSS colour (see Fig. 1).

The question of the zero-point requires some care. The primary photometric standard for the SDSS system is BD+17°4708, which has the colours $g - r = 0.29$ and $r - i = 0.1$; we therefore chose to set the SuperCOSMOS zero-point so that SDSS and SuperCOSMOS magnitudes are identical for galaxies with $g - r = 0.29$ (in the case of the $B_{\rm J}$ and $R_{\rm F}$ magnitudes) or $r - i = 0.1$ (in the case of $I_{\rm N}$ magnitudes). This means that our magnitudes are not precisely on an AB system, since this would require the magnitudes to match at $g - r = 0$ and $r - i = 0$. Very few galaxies are this blue, and we prefer to avoid an extrapolation away from the bulk of the calibrating data. But in any case, Fig. 1 suggests that any correction to exact AB would be small.

With this choice, the defining equations for the final SuperCOSMOS magnitudes are

$$\begin{aligned}
B_{\rm J} &= g - 0.024\Delta_{g-r} + 0.144\Delta^2_{g-r} \quad \text{(North)}\\
B_{\rm J} &= g - 0.015\Delta_{g-r} + 0.102\Delta^2_{g-r} \quad \text{(South)}\\
R_{\rm F} &= r - 0.133\Delta_{g-r} + 0.112\Delta^2_{g-r} \quad \text{(North)}\\
R_{\rm F} &= r - 0.085\Delta_{g-r} + 0.074\Delta^2_{g-r} \quad \text{(South)}\\
I_{\rm N} &= i - 0.081\Delta_{r-i} + 0.254\Delta^2_{r-i} \quad \text{(North)}\\
I_{\rm N} &= i - 0.151\Delta_{r-i} + 0.236\Delta^2_{r-i} \quad \text{(South)},
\end{aligned} \tag{1}$$

where $\Delta_{g-r} \equiv (g - r) - 0.29$ and $\Delta_{r-i} \equiv (r - i) - 0.1$. But for some purposes, it may be convenient to have the best-fitting linear relations, which are nearly as accurate:

$$\begin{aligned}
B_{\rm J} &= g - 0.078 + 0.134(g - r) \quad \text{(North)}\\
B_{\rm J} &= g - 0.058 + 0.102(g - r) \quad \text{(South)}\\
R_{\rm F} &= r + 0.012 - 0.054(g - r) \quad \text{(North)}\\
R_{\rm F} &= r - 0.002 - 0.022(g - r) \quad \text{(South)}\\
I_{\rm N} &= i + 0.008 - 0.024(r - i) \quad \text{(North)}\\
I_{\rm N} &= i + 0.022 - 0.092(r - i) \quad \text{(South)}.
\end{aligned} \tag{2}$$

These relations are intended to apply for $0 \lesssim g - r \lesssim 1.5$ and $0.1 \lesssim r - i \lesssim 0.7$. The rms accuracy for the quadratic fit is around 0.005 mag in $B_{\rm J}$ and $R_{\rm F}$ and 0.015 mag in $I_{\rm N}$; for the linear fit, the precision is a factor of 1.3 poorer.

For relation to older work, it may be useful to express these relations in terms of the Johnson–Cousins *UBVRI* system. Colour equations have been given that relate the SDSS data to the Johnson–Cousins system (Fukugita et al. 1996):

$$\begin{aligned}
B &= g + 0.217 + 0.419(g - r)\\
V &= g - 0.002 - 0.513(g - r)\\
R &= r - 0.155 - 0.089(g - r)\\
I &= i - 0.389 - 0.020(r - i) - 0.089(g - r)
\end{aligned} \tag{3}$$

(the last relation is only claimed to apply for $r - i < 0.9$, but this should apply for almost all galaxies here). The Fukugita et al. results were based on synthetic photometry, and so are vulnerable to uncertainties in filter profiles. A safer route is to use direct comparisons of photometry. This has been performed by Ivezić et al. (2007), comparing Johnson–Cousins *UBVRI* photometry from Stetson (2000, 2005) directly to SDSS magnitudes. Stetson's photometry is based on extensive CCD observations of the standards defined by Landolt (1992) using photomultipliers. Systematics in the relation of the two systems are at the 0.001 mag level, so that Stetson's photometry may be regarded as defining the Johnson–Cousins system in practice. The colour equations to SDSS are found to be non-linear, and Ivezić et al. give a set of transformations that involve a cubic in colour. But again, a set of linear transformations may be more

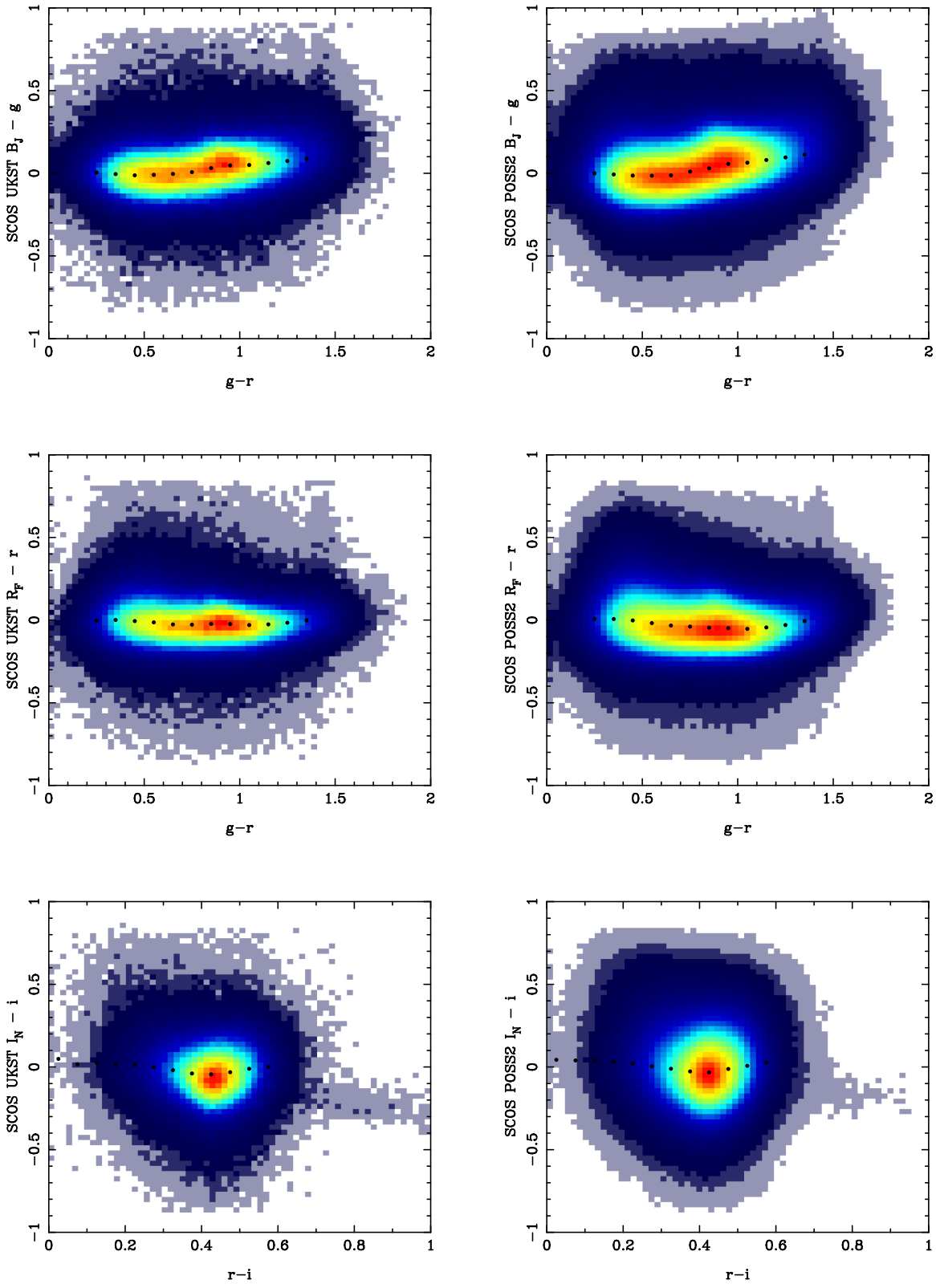

Figure 1. The colour relations with respect to SDSS photometry for the various SuperCOSMOS photographic bands, using limits of $B_J < 20$, $R_F < 19.5$, $I_N < 19$.

convenient, and the most useful form is an unpublished set of relations due to Lupton (2005):

$$B = g + 0.313(g - r) + 0.227$$
$$V = g - 0.578(g - r) - 0.004$$
$$R = r - 0.184(g - r) - 0.097$$
$$R = r - 0.294(r - i) - 0.144$$
$$I = i - 0.244(r - i) - 0.382. \quad (4)$$

We adopt these as the primary definitions of Johnson-Cousins magnitudes; in all cases, the scatter around these relations is close to 0.01 mag. In this case, and using our linear relations between SuperCOSMOS and SDSS, the SuperCOSMOS magnitudes may be expressed in terms of *BVRI* as

$$B_J = B - 0.259 - 0.201(B - V) \quad \text{(North)}$$
$$B_J = B - 0.230 - 0.237(B - V) \quad \text{(South)}$$
$$R_F = R + 0.089 + 0.215(V - R) \quad \text{(North)}$$
$$R_F = R + 0.070 + 0.267(V - R) \quad \text{(South)}$$
$$I_N = I + 0.335 + 0.231(R - I) \quad \text{(North)}$$
$$I_N = I + 0.356 + 0.160(R - I) \quad \text{(South)}. \quad (5)$$

3.2 Hemispheric colour-dependent offsets

From the above equations, it can be seen that some slight differences exist in the POSS2 and UKST photometric systems, despite the use of closely comparable observing materials. Ideally, one would like to eliminate these offsets. The most basic approach would be to adopt a typical colour and deduce an offset in zero-point. For objects that are not detected in all bands, there may be no other option. From Fig. 1, $g - r = 0.8$ and $r - i = 0.4$ are reasonable choices, yielding

$$B_J^{North} - B_J^{South} = +0.006$$
$$R_F^{North} - R_F^{South} = -0.015$$
$$I_N^{North} - I_N^{South} = +0.023. \quad (6)$$

In themselves, these offsets are insignificant. But if colours are available, one can do better. Because the I_N plates are less deep, this equalisation exercise is mainly of interest for the B_J and R_F magnitudes – since the deepest and most reliable catalogue arises from requiring detections in both of the shortest bands. In this case, one can use the following direct relations:

$$B_J^{North} - B_J^{South} = 0.03(B - R)^2 - 0.005(B - R)$$
$$R_F^{North} - R_F^{South} = 0.03(B - R)^2 - 0.06(B - R) + 0.015. \quad (7)$$

Here, $B - R$ denotes either photographic colour: the $N - S$ difference is unimportant in making this correction. For those applications where the I_N-plate depth is sufficient (e.g. 2MPZ; Bilicki et al. 2014), the corresponding correction is as follows:

$$I_N^{North} - I_N^{South} = 0.023(R - I)^2 + 0.06(R - I) - 0.01. \quad (8)$$

Note that the north–south corrections given in Bilicki et al. (2014) were incorrect, owing to an inadvertent swapping of hemispheres in their derivation.

3.3 Linearity and zero-points

Although SuperCOSMOS attempted to correct the measured photographic transmission to linear intensity, this is hard to carry out precisely. Each plate has its individual response to hypersensitization, and the finite scan resolution inevitably blends in saturated data from the centre of images. The issue of linearity therefore has to be treated empirically. For each plate, we fitted the following model:

$$m_{true} = a + b(m_{raw} + 2.5). \quad (9)$$

The offset in the second term is arbitrary, but was chosen to correspond to the typical SuperCOSMOS instrumental magnitudes for objects of reasonable S/N that were chosen for calibration – in order to remove any strong statistical coupling between the 'zero-point', a, and the 'linearity', b.

The resulting calibration parameters are shown in Figs 2 and 3, plotted against the progress of the surveys in time, via the unique number of each plate (as opposed to the field number). Some significant trends are visible, most notably the blue zero-point, which responds to secular variations in the brightness of the night sky. The redder zero-points and the linearity show smaller drifts with time. Generally, the linearity is unity within 5–10 per cent, and after correction the photometry may be assumed to be accurately linear – as may be verified by the residual plots shown in Fig. 4.

For plates where direct calibration is not available, the trends in Figs 2 and 3 yield an estimate of the calibration. The precision seems to be a little better for UKST than for POSS2: an rms of 2 per cent (POSS2)/1 per cent (UKST) in linearity and 0.16 mag (POSS2)/0.10 mag (UKST) in zero-point. For plates where direct calibration is lacking, this linearity estimate was adopted; for the zero-point, one can do better – as described below.

3.4 Corrections for vignetting and other field effects

The SDSS information also allows us to test for uniformity of the photometry within a given plate. It was expected that such variations would exist, if for no other reason than that geometrical vignetting within Schmidt telescopes results in less light reaching the edges of plates. We find that the mean offset in magnitude as a function of radius can be accurately described by

$$\Delta m = \frac{0.095(r - 2.^\circ25)^2}{1 + 0.025(r - 2.^\circ25)^2}, \quad (10)$$

and zero for $r < 2.^\circ25$. However, this behaviour is not universal, and individual plates exhibit radial trends that differ from this behaviour. After some experimentation, the following expression was found to be capable of capturing the empirical effect:

$$\Delta m = s\,(r - 1.^\circ5)^2, \quad (11)$$

and zero for $r < 1.^\circ5$. The 'slump coefficient' s needs to be determined for each plate; typical values are a few hundredths of a magnitude (generally positive, corresponding to loss of sensitivity at large r). The origin of such 'field effects' has been attributed to differential plate desensitization being affected by a varying gap between a flat filter and a curved focal surface. Whether or not this is the correct explanation, it does seem that the main effect is radially symmetric.

The only exception to the last statement is that small-scale variations were found perpendicular to the scan direction. These repeated over a distance corresponding to the SuperCOSMOS CCD and thus

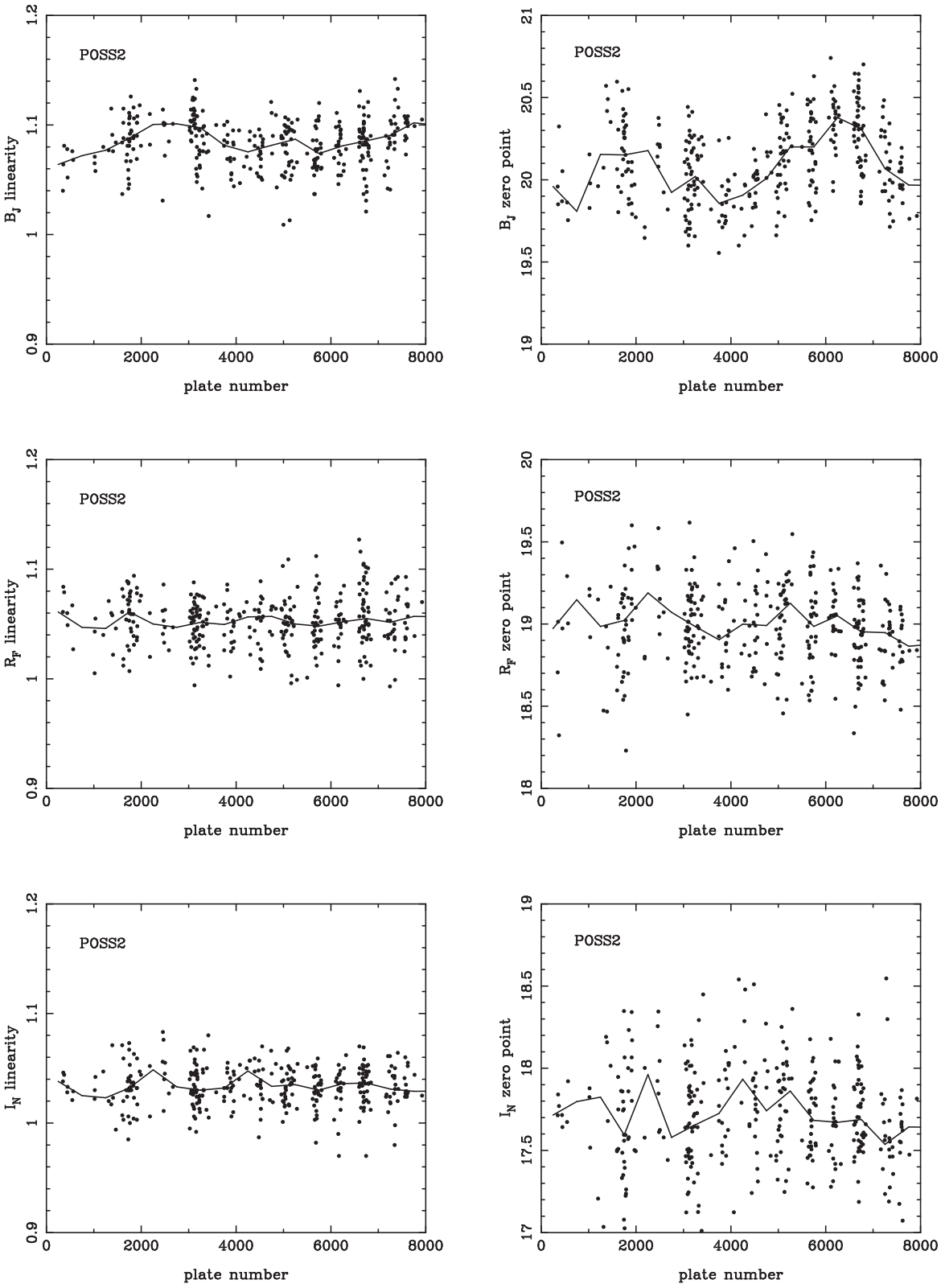

Figure 2. The derived calibration parameters for the POSS2 photographic plates, showing linearity and zero-point as a function of time (plate number), together with the running-average trends that were used as an initial estimate of the calibration when no SDSS photometry was available.

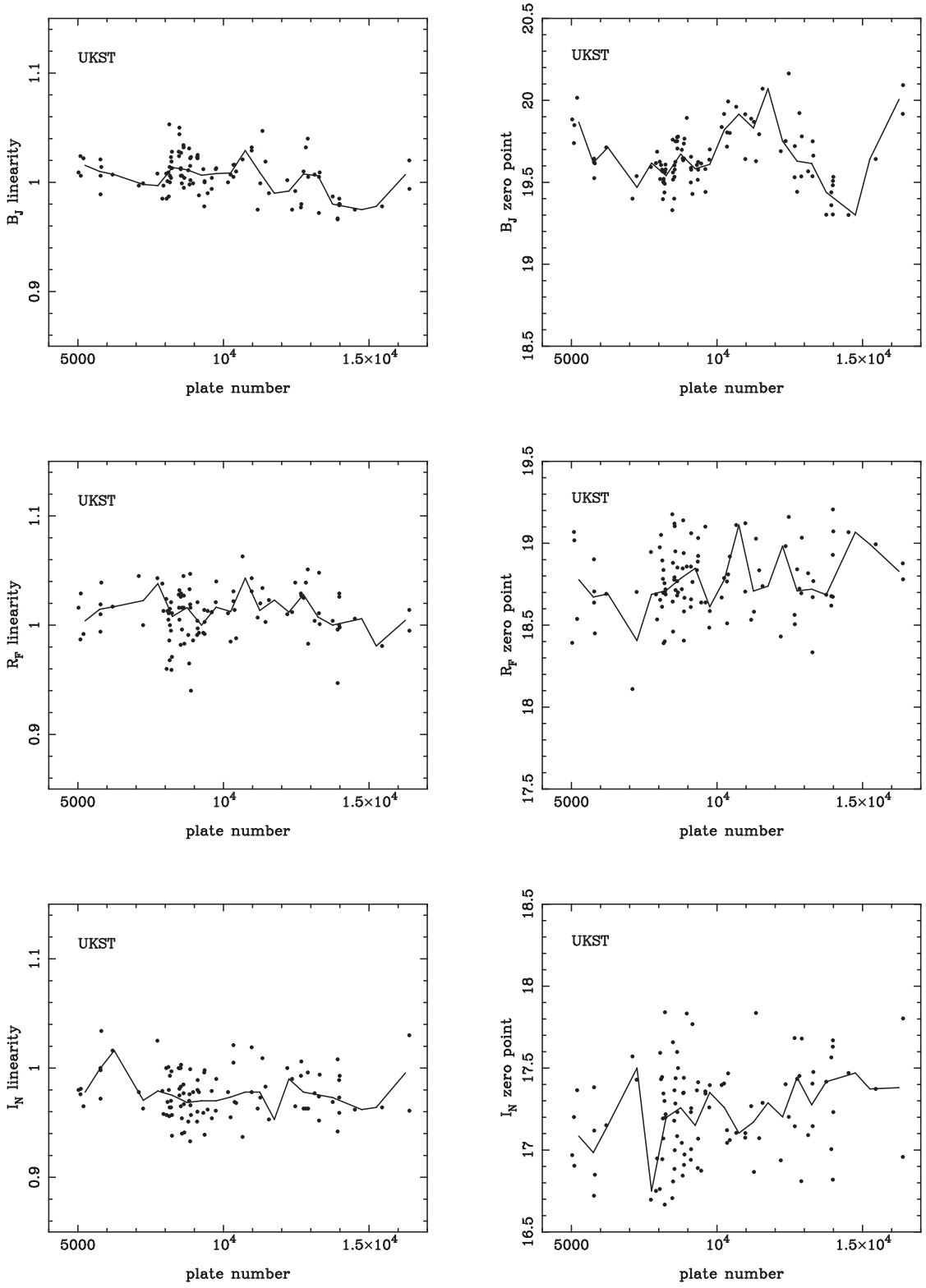

Figure 3. The southern equivalent of Fig. 2: derived calibration parameters for the UKST photographic plates, showing linearity and zero-point as a function of time (plate number), together with the running-average trends that were used as an initial estimate of the calibration when no SDSS photometry was available.

are plausibly attributable to uncorrected flat-field errors in the CCD. The mean effect is easily measured and is shown in Fig. 5. The correction is small enough that it was treated as being universal for all plates and wavebands.

3.5 Galactic extinction

For many science applications, we will want to select galaxies to a limit that is uniform with respect to magnitudes corrected for foreground extinction. This can be estimated using the colour

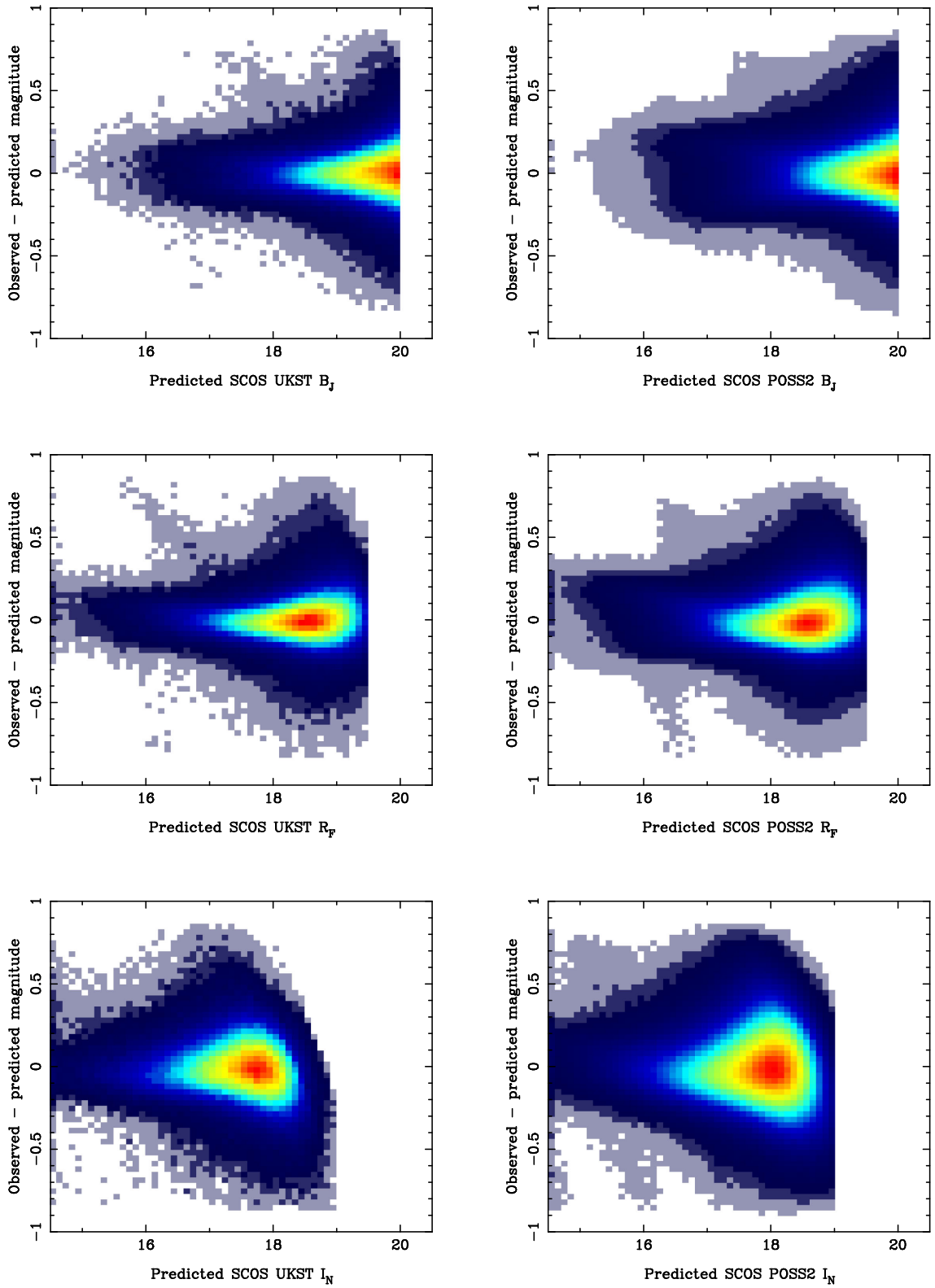

Figure 4. Magnitude residuals relative to SDSS, showing the uniformity and linearity of the SuperCOSMOS photometry after correction of individual plates. The principal selection for calibrating galaxies was $B_J < 20$.

equations given above. In order to use tabulated extinction coefficients for SDSS *ugriz*, we use the above linearized colour equations, expressed as a weighted linear combinations of magnitudes. These weights are then applied to the SDSS coefficients given in the $R_V = 3.1$ column from table 6 of Schlafly & Finkbeiner (2011). Note that this recalibration tends to yield coefficients that are substantially smaller than those given by Schlegel, Finkbeiner & Davis (1998), in part because the $E(B - V)$ estimates given by the

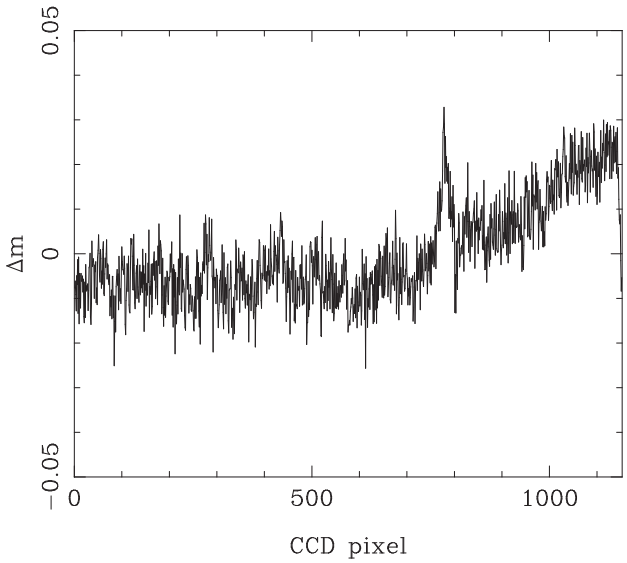

Figure 5. The effective flat-field correction required to account for additional sensitivity variations in the SuperCOSMOS CCD.

latter authors are generally too high by about 16 per cent. Schlafly & Finkbeiner's coefficients absorb this shift, and are designed to work with the published $E(B-V)$ estimates. We follow this convention for SuperCOSMOS, where the coefficients in $A = a\,E(B-V)_{\rm SFD}$ are 3.303, 2.285, 1.698 for (g, r, i), yielding

$$\begin{aligned} a(B_{\rm J}) &= 3.44 \quad \text{(North)} \\ a(B_{\rm J}) &= 3.41 \quad \text{(South)} \\ a(R_{\rm F}) &= 2.23 \quad \text{(North)} \\ a(R_{\rm F}) &= 2.26 \quad \text{(South)} \\ a(I_{\rm N}) &= 1.68 \quad \text{(North)} \\ a(I_{\rm N}) &= 1.64 \quad \text{(South)}. \end{aligned} \tag{12}$$

4 CALIBRATION WITH 2MASS

The SDSS calibration is essential for establishing the general systematic properties of the photographic data, but the whole point of the current exercise is that this calibration is unavailable over most of the sky. Although the general properties of plates seem predictable from their position in the observing sequence, this is not sufficient to tie down all properties (especially the zero-points) to an interesting precision.

The approach taken by the APM team was to bootstrap the calibration of plates using the 1.°25 overlap between plates. Although we incorporate this information, a more robust solution is possible through the near-IR 2MASS Extended Source Catalogue (XSC; Jarrett et al. 2000, 2003). We used the centroid coordinates (`sup_ra` and `sup_dec`) to find SuperCOSMOS counterparts of XSC sources and adopted the isophotal 2MASS photometry (`j_m_k20fe`, etc.). These magnitudes may have small systematic aperture corrections with redshift, but they are the most stable and therefore suitable for calibration via colour. The exact 2MASS XSC completeness limit is slightly soft. The design completeness was $K < 13.5$, but the number counts rise steeply until $K = 14$. We adopt a limit 0.5 mag brighter than the internal catalogue limit of 14.3: $K < 13.8$ (extinction-corrected). We also require corrected $J < 15.0$.

One can then look at the extinction-corrected optical-to-IR colours of galaxies on plates where SDSS calibration exists. Histograms of these colours are shown in Fig. 6. The breadth of these distributions naturally depends on wavelength, declining from an FWHM in colour of about 0.8 for $B_{\rm J} - J$ to a FWHM of 0.35 for $I_{\rm N} - J$. With ∼1000 2MASS galaxies per plate, that means that the average colour on each plate can be determined to a formal accuracy of 0.01 or better. We adjust the optical zero-point on a given plate to match the average optical–2MASS colour to the value known from plates with direct optical calibration. This should then in principle yield optical zero-points that are uniform to the all-sky photometric precision of 2MASS calibration, which is claimed to be 0.03 mag.

An earlier version of this procedure was used for refining the plate-dependent calibration of the APM magnitudes used by the 2dFGRS. But in addition to giving an estimate of the zero-point for each plate, we found in the present work that the number of 2MASS matches per plate is sufficient to permit an estimate of the slump coefficient, by determining the average colour in radial bins and fitting to equation (10). This means that relative offsets in plate zero-points can be determined for adjoining plates by matching galaxies in the overlap regions. The zero-points were then iterated to improve agreement between plates and their neighbours, allowing a maximum shift of 0.05 mag, in order to remain consistent with the expected absolute precision that should be delivered by the initial zero-points estimated from the 2MASS matching.

The plate zero-points used here were derived and applied to the SuperCOSMOS SSA data base at a time when the best available SDSS data set was DR6 (Adelman-McCarthy et al. 2008). The subsequent expansion of the SDSS would therefore now allow direct calibration for some plates where the zero-point was previously only inferred indirectly as above, but we have not updated the calibration in these cases. Rather, this paper serves as a documentation of the public SuperCOSMOS data set in the form in which it has been available and used for a number of years.

5 RELIABILITY OF THE CATALOGUE

In addition to calibration of the magnitudes, SDSS galaxy catalogues (treated as perfect for the current purposes) allow us to assess the limitations of the photographic data.

5.1 Trends with magnitude

Initially, we restrict attention to high Galactic latitudes ($|b| > 60^\circ$), in order to explore the limits of the data in the case where Galactic extinction and contamination is minimized. Fig. 7 shows the (sigma-clipped) dispersion in magnitude between SuperCOSMOS data and SDSS calibration, going deeper than the calibration constraints shown in Fig. 4. Effective magnitude limits are documented in Table 1. These are rather conservative, since Fig. 7 shows that the magnitude errors are never less than about 0.1 mag, even well above any plausible plate limit. This reflects a variety of limitations of the SuperCOSMOS data: emulsion measurement noise and the finite dynamic range of the SuperCOSMOS imaging system; imperfections in the photographic aperture corrections; and scatter in the colour relations with respect to SDSS. Thus, a fairer estimate of the random noise in the SuperCOSMOS magnitudes would probably subtract 0.1 in quadrature. In the face of these figures, there is a level of freedom in choosing limits to define a catalogue for practical use. We generally adopt the limits $B_{\rm J} < 21$, $R_{\rm F} < 19.5$: this takes the blue plates to their maximum depth, while requiring in addition a somewhat more precise detection in the red. This latter criterion

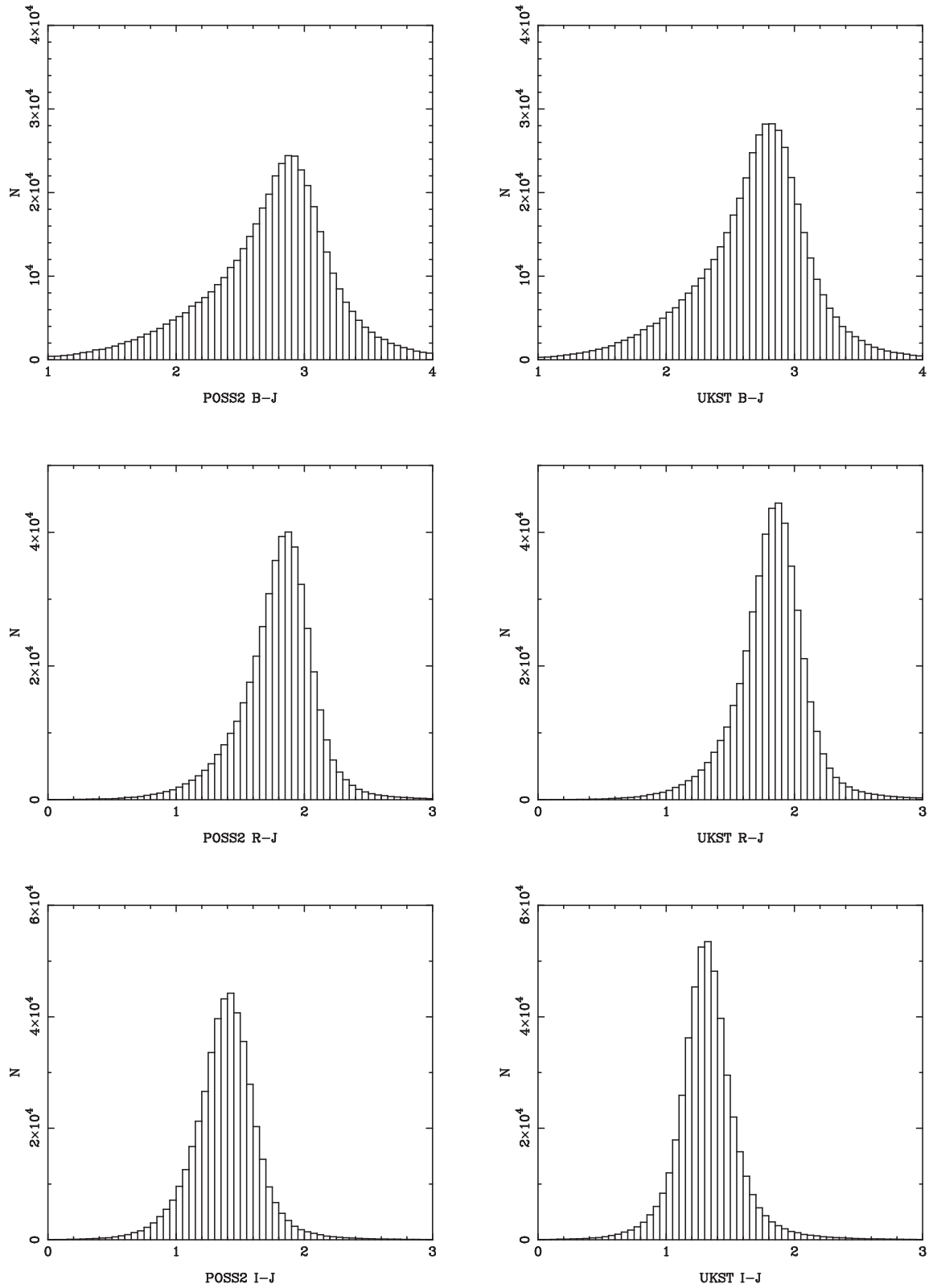

Figure 6. Histograms of extinction-corrected optical–2MASS colours after adjustment of plate zero-points. The average colour on each plate is used to yield an estimate of the zero-point for that plate, as well as limits on the radially dependent field effects.

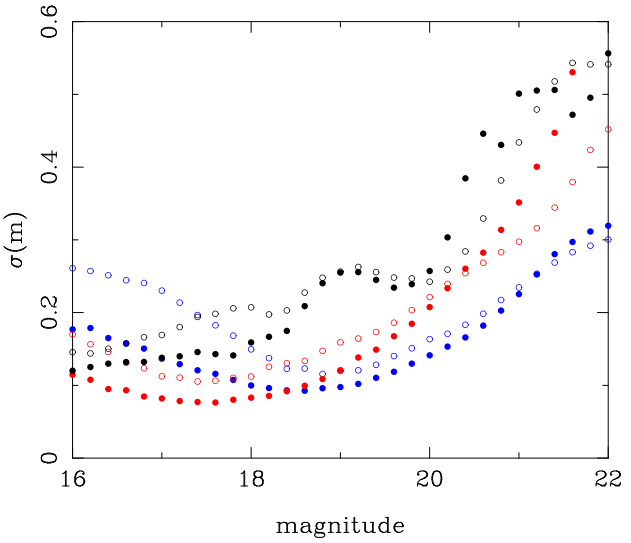

Figure 7. Magnitude errors as a function of depth, using a clipped rms. Filled points are UKST; open are POSS2. The different bands are B_J (blue), R_F (red) and I_N (black).

Table 1. SuperCOSMOS magnitude limits (see Fig. 7).

Band	5σ	4σ
UKST B_J	20.79	21.19
UKST R_F	19.95	20.30
UKST I_N	18.56	19.94
POSS2 B_J	20.26	21.17
POSS2 R_F	19.78	20.35
POSS2 I_N	18.38	18.90

permits more accurate colours, and is also effective at removing a range of artefacts that would exist in a single-plate catalogue (see Fig. 10).

5.2 Fidelity

The remaining question is how well the SuperCOSMOS catalogue approximates an ideal galaxy catalogue, subject to the magnitude errors shown in Fig. 7. We need to assess the purity and completeness of the catalogue: what fraction of our objects are actually galaxies, and what fraction of true galaxies are catalogued? This issue is discussed at some length in section 4.4 of Bilicki et al. (2016), with the conclusion that at high Galactic latitudes SuperCOSMOS attains a completeness of about 90 per cent, with 5 per cent stellar contamination. Inevitably, given the poorer image quality, compact galaxies are systematically lost from the sample and classified as stars. But this problem also afflicts digital surveys: in 2MASS, for instance, there is growing evidence that about half of the detected galaxies are in its *Point* Source Catalogue rather than the list of extended objects (Kovács & Szapudi 2015; Rahman, Ménard & Scranton 2016). Even SDSS misses about 2 per cent of galaxies in this way (Baldry et al. 2010).

For many purposes, the incompleteness and contamination figures documented above are tolerable. However, things degenerate at low Galactic latitudes. Fig. 8 shows the surface density as a function of sky position, where the main visible effect is a level of contamination rising to low Galactic latitudes and within the Magellanic Clouds, as stellar blending becomes more frequent. At the galactic poles, the surface density asymptotes to about 800 deg^{-2} for

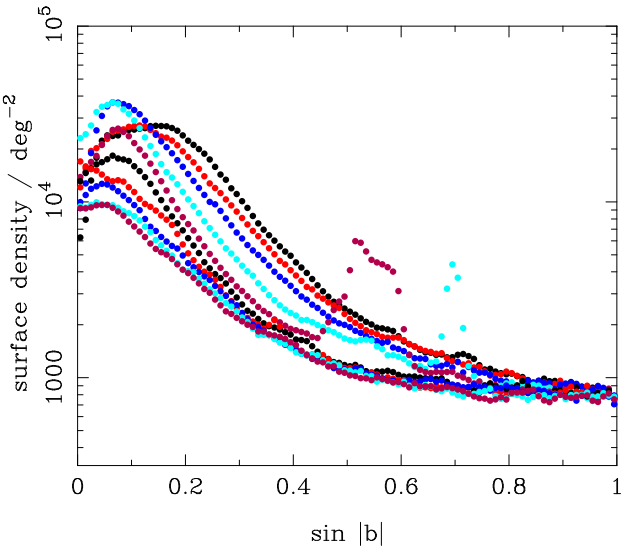

Figure 8. The total surface density of an extinction-corrected sample cut at $B_J < 21$ and $R_F < 19.5$, as a function of Galactic latitude, for various bins of Galactic longitude. With the exception of glitches arising from the Magellanic Clouds, the surface density of galaxy images is relatively constant for $|b| \gtrsim 30^\circ$, but stellar blend artefacts increase the apparent density by more than an order of magnitude at lower latitudes.

$B_J < 21$, $R_F < 19.5$, implying in principle 33 million real galaxies over the whole sky. But the density of artefacts rises steeply at low latitudes, typically doubling the apparent density at $|b| = 30^\circ$; at $|b| = 15^\circ$, the raw surface density is approximately 10 times higher than at the Galactic poles. Thus, the catalogue only has reasonable reliability over about half the sky (approximately 20 million galaxy candidates at $|b| > 30^\circ$, excluding the Magellanic Clouds, of which about 75 per cent are real).

Attempts were made to identify such spurious images based on discrepant extinction-corrected colours, but these were rather unsuccessful: the overlap of stellar and galaxy colours near the plate limits is too great. A significant reduction of this contamination is only possible when pairing with other wide-angle data sets. By using *WISE*, it was possible to generate a catalogue of 18.5 million galaxies with low contamination over 70 per cent of the sky (Bilicki et al. 2016).

5.3 Artefacts

Finally, it should be pointed out that the relatively good high-latitude performance of the SuperCOSMOS catalogue discussed above applies only to a sample paired up in B_J and R_F. If only a single waveband is used, the quality of the catalogue is significantly impaired. This point is illustrated in Fig. 9, which shows the sky density of galaxies selected to an (extinction-corrected) limit of $B_J < 21$. Based on considerations of S/N, this should be expected to be relatively clean; but a number of artefacts are apparent. The first is the offset in density between celestial hemispheres, owing to the slight difference in the photometric systems as detailed earlier. But there are also a number of other blemishes, which can be made more obvious if we pick out galaxies that fail to pair up with objects on the R_F plates. We choose for this purpose a cut at $R_F = 19.5$, and Fig. 10 plots just the objects with $B_J < 21$, $R_F > 19.5$. This reveals a variety of problems, ranging from large numbers of satellite trails, together with an inhomogeneous production of noise images in the

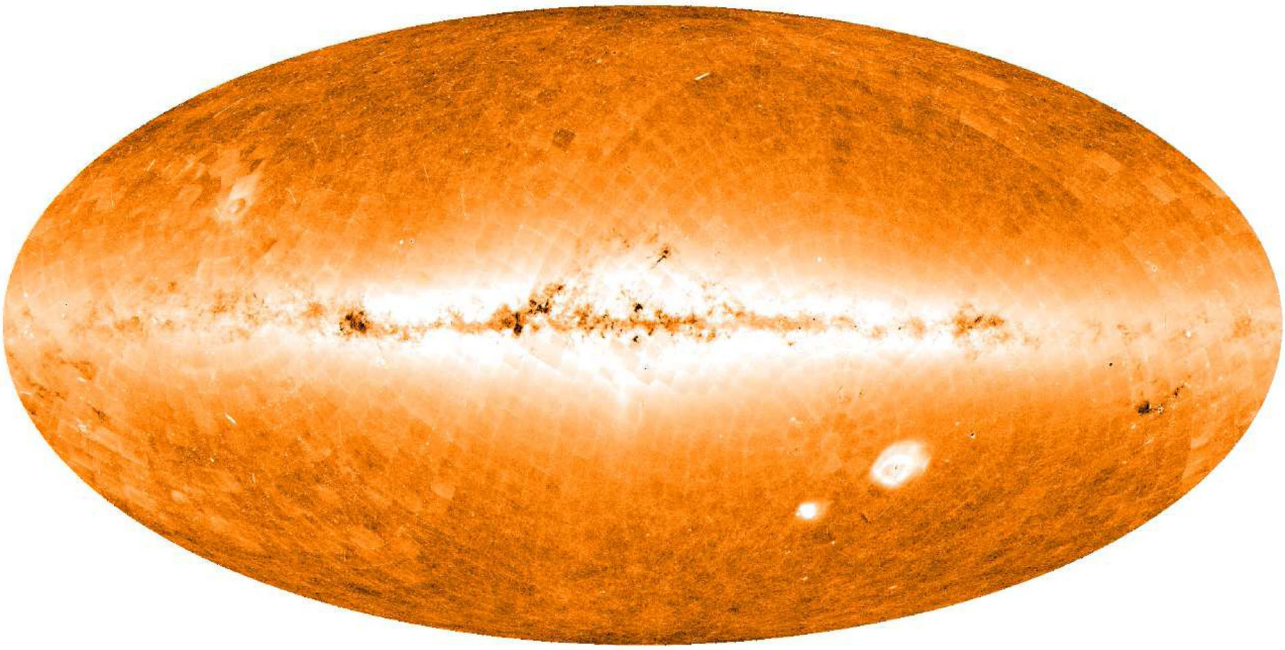

Figure 9. The sky distribution of the all-sky catalogue, to an extinction-corrected limit of $B_J < 21$, without a constraint on other wavebands. The Mollweide projection in Galactic coordinates shows surface density on a logarithmic scale, from 0.1 to 10 arcmin^{-2}. The hemispheric offset between the UKST and POSS2 photometric systems is apparent, and a number of artefacts can be seen. The latter are exposed in more detail in the next image.

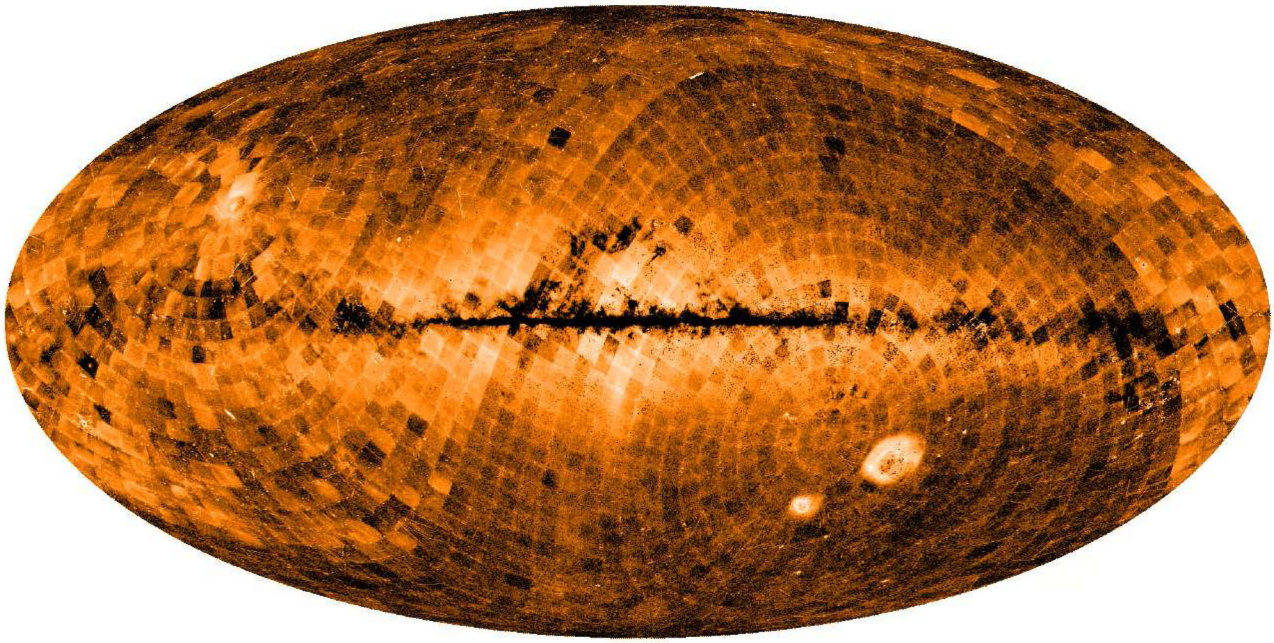

Figure 10. The sky distribution of the all-sky catalogue, to an extinction-corrected limit of $B_J < 21$, showing only objects fainter than $R_F = 19.5$. The Mollweide projection in Galactic coordinates shows surface density on a logarithmic scale, from 0.1 to 10 arcmin^{-2}. This image shows clearly the variety of artefacts that afflict a B_J-only selection, ranging from noise images in the less deep plates to satellite trails.

north, reflecting the greater variation in plate zero-points in that hemisphere (cf. Figs 2 and 3).

In contrast with these problems, the sky distribution of Fig. 11 is pleasingly clean. Here we require extinction-corrected $B_J < 21$ and $R_F < 19.5$, which also allows us to use the $B - R$ colour to correct the magnitudes to a single system (the northern one). This is the final data set used as input to the long-wavelength pairing exercises described in Bilicki et al. (2014, 2016), and it serves as the final extragalactic legacy of the decades of effort that were invested in Schmidt-plate sky surveys.

6 CONCLUSIONS

We have described the construction of a calibrated all-sky optical galaxy catalogue, based on SuperCOSMOS scans of Schmidt photographic plates from the UKST and POSS2 surveys.

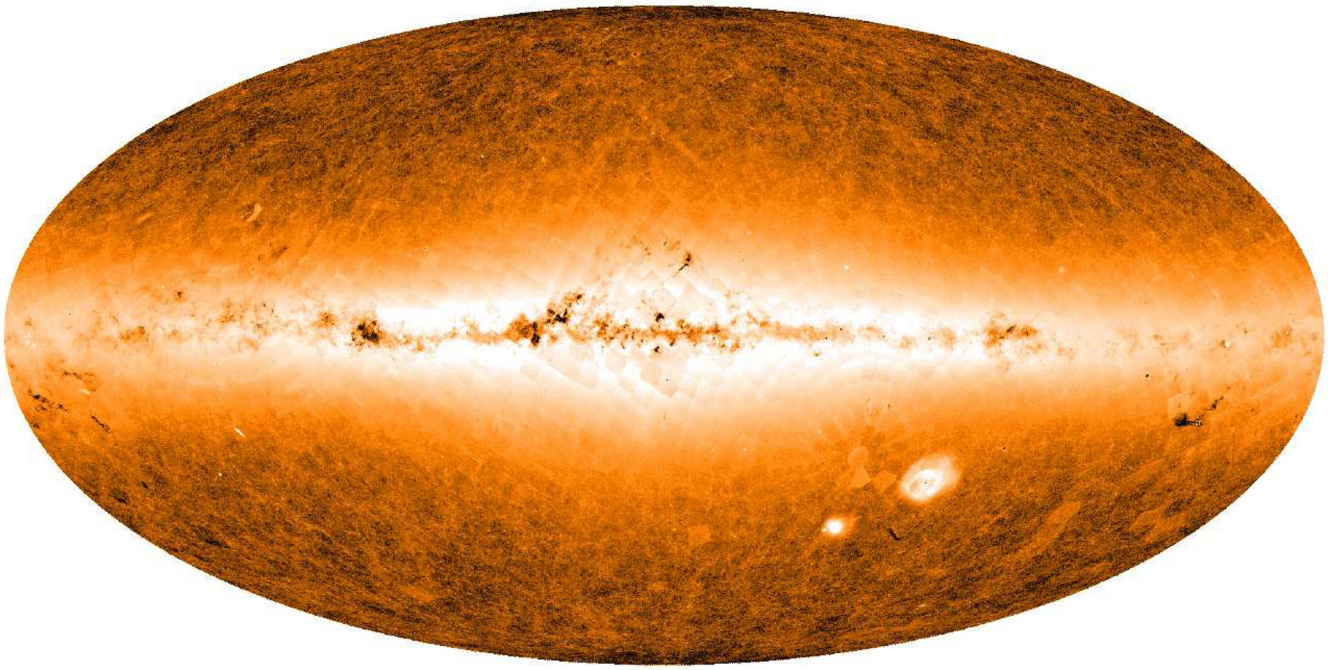

Figure 11. The sky distribution of the all-sky catalogue, to an extinction-corrected limit of $B_J < 21$, $R_F < 19.5$. The Mollweide projection in Galactic coordinates shows surface density on a logarithmic scale, from 0.1 to 10 arcmin^{-2}. The colours have been used to correct the southern magnitudes to the northern system, and there is now no visible hemispheric offset in surface density. Compared to selection in a single waveband, the sample is much cleaner, with the majority of artefacts removed. The main remaining issue is spurious galaxy images in regions of high stellar density, which cannot be removed without further information.

One immediate application of the catalogue has been to provide optical photometry of the 2MASS galaxy catalogue. This allows impressively accurate photometric redshifts to be estimated: an rms error of $\sigma_z \simeq 0.015$, probing the galaxy distribution out to $z \simeq 0.2$. A preliminary version of that work was used by Francis & Peacock (2010a,b) in a tomographic study of the foreground distortions of the cosmic microwave background due to the integrated Sachs–Wolfe effect. An in-depth study of 2MASS photometric redshifts from this data set was presented by Bilicki et al. (2014), accompanying a public release of the corresponding 2MPZ catalogue. Continuing this theme of combining the optical SuperCOSMOS data with longer wavelength catalogues, we have also constructed a catalogue of 18.5 million galaxies resulting from a cross-match with *WISE* data (limited at Vega $W1 < 17$ in its 3.4 μm band; see Wright et al. 2010), and this is described in detail in Bilicki et al. (2016). The photometric redshifts in this case are less precise than when 2MASS data are included, but are still impressive ($\sigma_z/(1 + z) \simeq 0.033$), and the catalogue covers the volume out to $z = 0.4$. In this way, we hope that the SuperCOSMOS galaxy catalogue will continue to be useful in cosmological research, while we await the happy day when SDSS-quality imaging becomes available over the whole sky. In principle, the raw material for such an all-sky digital galaxy catalogue should be available from Pan-STARRS (Kaiser et al. 2002) and SkyMapper (Keller et al. 2007).

The SuperCOSMOS all-sky galaxy catalogue described here is available at http://ssa.roe.ac.uk/allSky.html. The 2MPZ catalogue of photometric redshifts from matching SuperCOSMOS to 2MASS and *WISE* (Bilicki et al. 2014) can be found at http://ssa.roe.ac.uk/TWOMPZ.html. The matched *WISE*×SCOS catalogue described by Bilicki et al. (2016) can be found at http://ssa.roe.ac.uk/WISExSCOS.html.

ACKNOWLEDGEMENTS

This research has made use of data obtained from the SuperCOSMOS Science Archive, consisting of scanned survey plates from the UK Schmidt Telescope and the Palomar Observatory Sky Survey (POSS-II). This archive is prepared and hosted by the Wide Field Astronomy Unit, Institute for Astronomy, University of Edinburgh, which is funded by the UK Science and Technology Facilities Council.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation and the US Department of Energy Office of Science. The SDSS-III website is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington and Yale University.

This publication makes use of data products from the 2MASS, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation; the NASA/IPAC

Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration and of the NASA/IPAC Infrared Science Archive, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

We thank Robert Lupton for helpful correspondence about SDSS colour equations.

JAP was supported by a PPARC Senior Research Fellowship during earlier parts of this work. He is currently supported by ERC grant number 670193. MB is supported by the Netherlands Organization for Scientific Research, NWO, through grant number 614.001.451, and through FP7 grant number 279396 from the European Research Council.

REFERENCES

Adelman-McCarthy J. K. et al., 2008, ApJS, 175, 297
Alam S. et al., 2015, ApJS, 219, 12
Astier P., Pain R., 2012, C. R. Phys., 13, 521
Aubourg É et al., 2015, Phys. Rev. D, 92, 123516
Baldry I. K. et al., 2010, MNRAS, 404, 86
Bilicki M., Jarrett T. H., Peacock J. A., Cluver M. E., Steward L., 2014, ApJS, 210, 9
Bilicki M. et al., 2016, ApJS, 225, 5
Blair M., Gilmore G., 1982, PASP, 94, 742
Colless M. et al., 2001, MNRAS, 328, 1039
Colless M. et al., 2003, preprint (astro-ph/0306581)
Doyle M. T. et al., 2005, MNRAS, 361, 34
Efstathiou G., Sutherland W. J., Maddox S. J., 1990, Nature, 348, 705
Efstathiou G. et al., 2002, MNRAS, 330, L29
Francis C. L., Peacock J. A., 2010a, MNRAS, 406, 2
Francis C. L., Peacock J. A., 2010b, MNRAS, 406, 14
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Hambly N. C. et al., 2001a, MNRAS, 326, 1279
Hambly N. C., Irwin M. J., MacGillivray H. T., 2001b, MNRAS, 326, 1295
Hambly N. C., Davenhall A. C., Irwin M. J., MacGillivray H. T., 2001c, MNRAS, 326, 1315
Ivezić Ž. et al., 2007, in Sterken C., ed., ASP Conf. Ser. Vol. 364, The Future of Photometric, Spectrophotometric and Polarimetric Standardization. Astron. Soc. Pac., San Francisco, p. 165
Jarrett T. H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J. P., 2000, AJ, 119, 2498
Jarrett T. H., Chester T., Cutri R., Schneider S. E., Huchra J. P., 2003, AJ, 125, 525
Jones D. H. et al., 2009, MNRAS, 399, 683
Kaiser N. et al., 2002, Proc. SPIE, 4836, 154
Keller S. C. et al., 2007, PASA, 24, 1
Kovács A., Szapudi I., 2015, MNRAS, 448, 1305
Kurcz A., Bilicki M., Solarz A., Krupa M., Pollo A., Małek K., 2016, A&A, 592, A25
Landolt A. U., 1992, AJ, 104, 340
Lupton R., 2005. Available at: https://www.sdss3.org/dr10/algorithms/sdssUBVRITransform.php
Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990a, MNRAS, 243, 692
Maddox S. J., Efstathiou G., Sutherland W. J., 1990b, MNRAS, 246, 433
Morrissey P. et al., 2007, ApJS, 173, 682
Murphy T. et al., 2010, MNRAS, 402, 2403
Planck Collaboration XIII 2015, preprint (arXiv:1502.01589)
Rahman M., Ménard B., Scranton R., 2016, MNRAS, 457, 3912
Reid I. N. et al., 1991, PASP, 103, 661
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Stetson P., 2000, PASP, 112, 925
Stetson P., 2005, PASP, 117, 563
Wright E. L. et al., 2010, AJ, 140, 1868

This paper has been typeset from a TEX/LATEX file prepared by the author.